DESCRIPTION

Novel N-Acetylgalactosamine Transferases and Nucleic Acids Encoding the Same

Technical Field

The present invention relates to novel enzymes having the activity of transferring N-acetylgalactosamine to N-acetylglucosamine via a β1-4 linkage and nucleic acids encoding the same, as well as to nucleic acids for assaying said nucleic acids.

Background Art

In various kinds of organisms, structures having a linkage of disaccharide of N-acetylgalactosamine-N-acetylglucosamine have been found in oligosaccharides of glycoproteins and glycolipids [see References 1 and 2]. In humans, this disaccharide structure is known as a β1-4 linkage (GalNAcβ1-4GlcNAc), and is found only in N-glycans [see Reference 3]. Methods for obtaining human-type oligosaccharides including said structure are limited to methods using complicated chemical synthesis and methods obtaining the oligosaccharides from natural proteins. Further, the above disaccharide structure includes in vivo a galactose substituted for a N-acetylgalactosamine. Therefore, it is a lengthy, laborious process to obtain oligosaccharides having the target disaccharide structure.

Prior to the present application, the inventors identified ppGalNAc-T10, -T11, -T12, -T13, -T14, -T15, -T16, -T17, CSGalNAc-T1, and -T2 as enzymes having an activity of

transferring N-acetylgalactosamine to glucuronic acids and polypeptides, and further, they clarified the structures of these genes. Already known are at least 22 N-acetylgalactosamine transferases that have the activity of transferring N-acetylgalactosamine (Table 1), and each of the transferases have different specificities of acceptor substrates.

Table 1 N-acetylgalactosamine transferase and the substrate specificity

Formal Name	Abbreviation	Origin	Substrate specificity	References
UDP-GalNAc:polypeptide N-acetylgalactosaminyl transferase I	ppGalNAc-T1	human	Ser/Thr	White, T. etc (1995)
UDP-GalNAc:polypeptide N-acetylgalactosaminyl transferase II	ppGalNAc-T2	human	Ser/Thr	White, T. etc (1995)
UDP-GalNAc:polypeptide N-acetylgalactosaminyl transferase III	ppGalNAc-T3	human	Ser/Thr	Bennet, E. P. etc (1996)
UDP-GalNAc:polypeptide N-acetylgalactosaminyl transferase IV	ppGalNAc-T4	human	Ser/Thr	Bennet, E. P. etc (1998)
UDP-GalNAc:polypeptide N-acetylgalactosaminyl transferase VI	ppGalNAc-T6	human	Ser/Thr	Bennet, E. P. etc (1999) (1)
UDP-GalNAc:polypeptide N-acetylgalactosaminyl transferase VII	ppGalNAc-T7	human	Ser/Thr	Bennet, E. P. etc (1999) (2)
UDP-GalNAc:polypeptide N-acetylgalactosaminyl transferase VIII	ppGalNAc-T8	human	Ser/Thr	White, K. E. etc (2000)
UDP-GalNAc:polypeptide N-acetylgalactosaminyl transferase IX	ppGalNAc-T9	human	Ser/Thr	Toba, S. etc (2000)
UDP-GalNAc:polypeptide N-acetylgalactosaminyl transferase X	ppGalNAc-T10	human	Ser/Thr	JP No. 2001-401455 (unpublished)
UDP-GalNAc:polypeptide N-acetylgalactosaminyl transferase XI	ppGalNAc-T11	human	Ser/Thr	JP No. 2001-401507 (unpublished)
UDP-GalNAc:polypeptide N-acetylgalactosaminyl transferase XII	ppGalNAc-T12	human	Ser/Thr	JP No. 2001-401507 (unpublished)
UDP-GalNAc:polypeptide N-acetylgalactosaminyl transferase XIII	ppGalNAc-T13	human	Ser/Thr	JP No. 2001-401507 (unpublished)
UDP-GalNAc:polypeptide N-acetylgalactosaminyl transferase XIV	ppGalNAc-T14	human	Ser/Thr	Guo, J. M. etc (2002)
UDP-GalNAc:polypeptide N-acetylgalactosaminyl transferase XV	ppGalNAc-T15	human	Ser/Thr	JP No. 2001-401507 (unpublished)
UDP-GalNAc:polypeptide N-acetylgalactosaminyl transferase XVI	ppGalNAc-T16	human	Ser/Thr	JP No. 2001-401507 (unpublished)
UDP-GalNAc:polypeptide N-acetylgalactosaminyl transferase XVII	ppGalNAc-T17	human	Ser/Thr	JP No. 2001-401507 (unpublished)
β1, 4-N-acetylglactosamine transferase	β4GalNAcT	human	GM3, GD3, LacCer	Nagata, Y. etc (1992)
UDP-GalNAc:H-α1, 3-N-acetylgalactosamine transferase	Hist blood A group transferase	human	Fucα1, 2Galβ1-R	Yamamoto, F. etc (1990)
UDP-GalNAc:globoside α1, 3-N-acetylgalactosamine transferase I	formalin glycolipid synthase	human	GalNAcβ1-3Galα1 -4Galβ1-3Glc-Cer	Xu, H. etc (1999)
Chondroitin sulfate N-acetylglactosamin transferase I	CSGalNAc-T1	human	GlcA	JP No. 2002-129156 (unpublished)
Chondroitin sulfate N-acetylglactosamin transferase II	CSGalNAc-T2	human	GlcA	JP No. 2002-24202 (unpublished)

Disclosure of Invention

Isolation of an enzyme having the activity of transferring N-acetylgalactosamine to N-acetylglucosamine via a β1-4 linkage and an explanation of the structure of its gene enable the production of said enzyme or the like through genetic engineering techniques, and the diagnosis of diseases on the basis of said gene or the like. However, such an enzyme has not been isolated/purified yet and there is no key to isolating such an enzyme and identifying its gene. Therefore, no antibody against such an enzyme has been prepared.

Therefore, the present invention provides a protein having an activity of transferring N-acetylgalactosamine to N-acetylglucosamine via a β1-4 linkage and nucleic acids for encoding the same. The present invention also provides a cell introduced with a recombinant vector expressing said nucleic acids in a host cell and said nucleic acids, and expressing said nucleic acids and said proteins. Further, said protein expressed can be used for producing an antibody. Therefore, the present invention also provides a method for producing said protein. Further, the expressed protein and said antibody to the protein can be applied to immnohistochemical staining, and immunoassay of RIA and EIA and the like. Moreover, the present invention provides an analytical nucleic acid for assaying the above nucleic acid of the present invention.

As described above, the objective enzymes have not yet been identified, and therefore, the partial sequence of the amino acids cannot be informed. In general, it is

difficult to isolate and purify proteins which are included in only a very small quantity in cells. Therefore, it is supposed that it is not easy to isolate enzymes which have so far not been isolated from cells. Thereat, the inventors tried to isolate and purify target enzymes, by making a region of which identity is thought to be high into a target, which may have the homologous sequence in nucleic acid sequences of genes between a objective enzyme and various kinds of enzymes having relatively similar activity. Specifically, the inventors first searched nucleic acid sequences of publicly-known β1,4-galactose transferases, and identified homologous regions. Second, primers were designed based on these homologous regions, and a full-length open reading flam was identified from cDNA library by 5' RACE (rapid amplification of cDNA ends) method. Further, the inventors succeeded in cloning a gene of said enzyme by PCR, and completed the present invention by determining nucleic acid sequences thereof and putative amino acid sequences.

The present invention provides a protein having the activity of transferring N-acetylgalactosamine and nucleic acid encoding the same, and thereby assists in satisfying these various requirements in the art.

Namely, the present invention provides a mammal protein having the activity of transferring N-acetylgalactosamine to N-acetylglucosamine via a β1-4 linkage.

The human protein of the present invention has, typically, amino acid sequence of SEQ ID NO: 1 or 3, which

is presumed from nucleic acid sequence of SEQ ID NO: 2 or 4.

The mouse protein of the present invention has amino acid sequence of SEQ ID NO: 26 or 28, which is presumed from nucleic acid sequence of SEQ ID NO: 27 or 29.

The present invention includes not only the protein having the amino acid sequence which is selected from a group consisting of SEQ ID NOs: 1, 3, 26 and 28 but also proteins having an identity of 50 % or more to said sequence. The present invention includes proteins having said amino acid sequence, wherein one or more amino acids are substituted or deleted, or one or more amino acids are inserted or added.

The proteins of the present invention have amino acid sequences which have an identity of 60 % or more, preferably 70 % or more, more preferably 80 % or more, still more preferably 90 %, and most preferably 95 % to the amino acid sequence which is selected from a group consisting of SEQ ID NOs: 1, 3, 26 and 28.

The present invention provides nucleic acids encoding the protein of the present invention.

The nucleic acids of the present invention have, typically, the nucleic acid sequence which is selected from a group consisting of SEQ ID NOs: 2, 4, 27 and 29, nucleic acid sequences in which one or more nucleic acids are substituted, deleted, inserted and/or added to the above nucleic acid sequence, or a nucleic acid sequence which hybridizes with said nucleic acid sequence under stringent conditions, and which includes the nucleic acids complementary to the above sequences. In one embodiment,

the present invention includes, but is not limited to, nucleic acids having the nucleic acid sequence represented by nucleotides 1-3120 of the nucleic acid sequence shown in SEQ ID NO: 2, nucleotides 1-2997 of the nucleic acid sequence shown in SEQ ID No: 4, nucleotides 1-3105 of the nucleic acid sequence shown in SEQ ID NO: 27, nucleotides 1-2961 of the nucleic acid sequence shown in SEQ ID No: 29.

The present invention provides a recombinant vector containing the nucleic acids of the present invention.

The present invention provides the transformants obtained by introducing the recombinant vector of the present invention into host cells.

The present invention provides an analytical nucleic acid which hybridizes to the nucleic acids encoding the protein of the present invention under stringent conditions. The analytical nucleic acid preferably has the sequence shown in any one of SEQ ID NOs: 20, 21, 23 and 24 in the case of using the analytical nucleic acid of the present invention as a probe for assaying the nucleic acids encoding said protein. Further, the analytical nucleic acid of the present invention can be used as a cancer marker.

The present invention provides an assay kit comprising the analytical nucleic acid which hybridizes to the nucleic acid of the present invention.

The present invention provides the isolated antibody binding to the protein of the present invention or the monoclonal antibody thereof.

Further, the present invention provides a method for determining a canceration of biological sample which

comprises a step of quantifying the protein or the nucleic acid of the present invention in the biological sample.

Brief Description of Drawings

Fig. 1 is a graph showing the quantitative analysis of expression level of NGalNAc-T1 or NGalNAc-T2 gene in various human tissues by the real time PCR. The axis of ordinates represents a relative ratio of expression level of NGalNAc-T1 or NGalNAc-T2 gene to that of a control glyceraldehyde-3-phsopate dehydrogenase (GAPDH) gene. The expressions of NGalNAc-T1 and NGalNAc-T2 gene are represented as a black bar and a white bar, respectively.

Fig. 2 is a graph showing the quantitative analysis of expression level of NGalNAc-T1 (panel A) or NGalNAc-T2 (panel B) gene in human lung cancerous tissue and normal tissue by the real time PCR. The axis of ordinates represents a relative ratio of expression level of NGalNAc-T1 or NGalNAc-T2 gene to that of a control human β-actin gene. The axis of abscissas represents numbers relating to each patient. The normal tissue and the cancerous tissue are represented as a white bar and a black bar, respectively.

Fig. 3 shows LacdiNAc synthesizing activity of NGalNAc-T2 toward asialo/agalacto-fetal calf fetuin. The asialo/agalacto-FCF appears as approximately 55 and 60 kDa band (lane 1). The NGalNAc-T2 effectively transfers GalNAc to asialo/agalacto-FCF (lane 5). The band mostly disappeared by GPF treatment (lane 6).

Fig. 4 shows an analysis of N-glycan structures of glycodelin from NGalNAc-T1 and NGalNAc-T2 gene transfected

CHO cells. The non-reducing terminal GalNAc is detected only when NGalNAc-T1 or NGalNAc-T2 gene is co-transfected with glycodelin gene.

Fig. 5 shows one-dimensional ^{1}H NMR spectrum of the structure of GalNAcb1-4GlcNAc-O-Bz produced by NGalNAc-T2.

Fig. 6 shows two-dimensional ^{1}H NMR spectrum of the structure of GalNAcb1-4GlcNAc-O-Bz produced by NGalNAc-T2.

Detailed Description of the Invention

In order to explain the present invention, a preferable embodiments for carrying out the invention are described in detail below.

(1) Proteins

The nucleic acid encoding the human protein of the present invention cloned by the method described in detail in the examples below has the nucleotide sequence shown in SEQ ID NO: 2 or 4 in the Sequence Listing under which a deduced amino acid sequence encoded thereby is also shown. In addition, SEQ ID NO: 1 or 3 shows only said amino acid sequence.

The proteins (hereinafter, denominated "NGalNAc-T1" and "NGalNAc-T2") of the present invention obtained in the examples below are enzymes having the properties listed below. In addition, each property of the proteins of the present invention and the method for determining the activity thereof are described in detail in the examples below.

Activity: Transferring N-acetylgalactosamine to N-acetylglucosamine via a β1-4 linkage. The catalytic reaction is represented by
the reaction formula:

UDP-N-acetyl-D-galactosamine + N-acetyl-D-glucosamine-R -> UDP + N-acetyl-D-galactosaminyl-N-acetyl-D-glucosamine-R
(UDP-GalNAc + GlcNAc-R -> UDP + GalNAc-GlcNAc-R)

Specific substrate: N-acetyl-glucosamine such as N-acetylglucosamine β1-3-R (R is a residue of which hydroxyl group of mannose and p-nitrophenol and the like binds via an ether linkage).

In a preferable embodiment, the proteins of the present invention have at least one of the following properties, preferably these properties:

(A) Specificity of acceptor substrates

(a) When O-linked oligosaccharides are used as an acceptor substrate, said proteins have the activity of transferring N-acetylgalactosamine to GlcNAcβ1-6(Galβ1-3)GalNAcα-pNp (hereinafter, "core2-pNp"), GlcNAcβ1-3GalNAcα-pNp (hereinafter, "core3-pNp"), GlcNAcβ1-6GalNAcα-pNp (hereinafter, "core6-pNp") via a β1-4 linkage, wherein the abbreviations used are: GlcNAc, N-acetylglucosamine; GalNAc, N-acetylgalactosamine; Gal, galactose; pNp, p-nitrophenyl. Preferably, said proteins have the transferring activity to core6-pNp.

(b) When N-linked oligosaccharides are used as an acceptor substrate, said proteins have the activity of

transferring N-acetylgalactosamine to GlcNAc at the non-reducing end of said oligosaccharides via a β1-4 linkage, provided that said activity reduces when said oligosaccharides have the following properties:

(i) having fucose (Fuc) residues in the structure of said oligosaccharides; and

(ii) having one or more branched chains wherein GalNAc residues bind to GlcNAc residues at the non-reducing end.

(B) Optimum pH in enzymatic activity

The activity tends to be higher in pH 6.5 of MES (2-morpholineethanesulfonic acid) buffer. In HEPES ([4-(2-hydroxyethyl)-1-piperazinyl]ethanesulfonic acid) buffer, the activity tends to be higher in pH 6.75 for NGalNAc-T1 and pH 7.4 for NGalNAc-T2.

(C) Requirement of divalent ions

In NGalNAc-T1, the activity tends to be higher in the MES buffer including at least Mn^{2+}, or Cu^{2+}, preferably Mn^{2+}. In NGalNAc-T2, the activity tends to be higher in the MES buffer including Mg^{2+}, Mn^{2+}, or Co^{2+}, preferably Mg^{2+}.

The nucleic acid encoding the mouse protein of the present invention also has the nucleotide sequence shown in SEQ ID NO: 27 or 29 in the Sequence Listing under which a deduced amino acid sequence encoded thereby is also shown. In addition, SEQ ID NO: 1 or 3 shows only said amino acid sequence. The proteins (hereinafter, denominated "mNGalNAc-T1" and "mNGalNAc-T2") of the present invention are enzymes having the above properties.

The present invention provides a protein having an activity for transferring N-acetylgalactosamine to N-acetylglucosamine via a β1-4 linkage. So far as the proteins of the present invention have the properties described herein, the origins thereof and the method for producing them and the like are not limited. Namely, the proteins of the present invention include, for example, native proteins, proteins expressed from recombinant DNA using genetic engineering techniques, and chemically synthesized proteins.

The protein of the present invention has typically an amino acid sequence consisting of 1039 amino acids shown in SEQ ID NO: 1, 998 amino acids shown in SEQ ID NO: 3, 1034 amino acids shown in SEQ ID NO: 26, or 986 amino acids shown in SEQ ID NO: 28. However, it is well-known that in native proteins, there are mutant proteins having one or more variants of amino acids, depending on a mutation of gene based on various species of organisms which produce the proteins, and various ecotypes, or a presence of very similar isozymes or the like. In addition, the term "mutant protein(s)" used herein means proteins and the like having a variant of said amino acid sequence, wherein one or more amino acids are substituted or deleted, or one or more amino acids are inserted or added in the amino acid sequence of SEQ ID NO: 1, 3, 26 or 28, and having the activity of transferring N-acetylgalactosamine to N-acetylglucosamine via a β1-4 linkage. The expression "one or more" here preferably means 1-300, more preferably 1-100, and most

preferably 1-50. Generally, in the instance that amino acids are substituted by site-specific variation, the number of amino acids that can be substituted to the extent that the activity of the original protein can be retained is preferably 1-10.

Proteins of the present invention have the amino acid sequences of SEQ ID NO: 1 or 3 and SEQ ID NO: 2 or 4 (lower), or amino acid sequences of SEQ ID NO: 26 or 28 and SEQ ID NO: 27 or 29 (lower) based on the premise of nucleotide sequences of the cloned nucleic acids, but are not exclusively limited to the proteins having these sequences, and are intended to include all homologous proteins having the characteristics described herein. The identity is at least 50 % or more, preferably 60 %, more preferably 70 % or more, even more preferably 80 % or more, still more preferably 90 % or more, and most preferably 95 % or more.

As used herein, the percentage identity of amino acid sequences can be determined by comparison with sequence information using, for example, the BLAST program described by Altschul et al. (Nucl. Acids. Res. 25, pp. 3389-3402, 1997) or the FASTA program described by Pearson et al. (Proc. Natl. Acad. Sci. USA, pp. 2444-2448, 1988). These programs are available from the website of National Center for Biotechnology Information (NCBI) or DNA Data Bank of Japan (DDBJ) on the Internet. Various conditions (parameters) for homology searches with each program are described in detail on the site, and searches are normally performed with default values though some settings may be appropriately changed. Other programs used by those skilled in the art of

sequence comparison may also be used.

Generally, a modified protein containing a change from one amino acid to another amino acid having similar properties (such as a change from a hydrophobic amino acid to another hydrophobic amino acid, a change from a hydrophilic amino acid to another hydrophilic amino acid, a change from an acidic amino acid to another acidic amino acid or a change from a basic amino acid to another basic amino acid) often has similar properties to those of the original protein. Methods for preparing such a recombinant protein having a desired variation using genetic engineering techniques are well known to those skilled in the art and such modified proteins are also included in the scope of the present invention.

Proteins of the present invention can be obtained in bulk by, for example, introducing and expressing the DNA sequence of SEQ ID NO: 2, 4, 27 or 29 representing a nucleic acid of the present invention in E. coli, yeast, insect or animal cells using an expression vector capable of being amplified in each host, as described in the examples below.

When the identity search of the protein of the present invention is performed using GENETYX (Genetyx Co.), the NGalNAc-T1 has 47.2 % identity to NGalNAc-T2, 84.3 % identity to mNGalNAc-T1, and 47.4 % identity to mNGalNAc-T2. The NGalNAc-T2 has 46.5 % identity to mNGalNAc-T1, and 82.6 % identity to mNGalNAc-T2. The mNGalNAc-T1 has 46.3 % identity to mNGalNAc-T2.

The NGalNAc-T1 has 26.1 % identity in 226 amino acids of C terminus to CSGalNAc-T1, while the NGalNAc-T2 has

21.6 % identity in 431 amino acids of C terminus to CSGalNAc-T1 and 25.0 % identity in 224 amino acids of C terminus to CSGalNAc-T2.

Further, the NGalNAc-T1 has 19.3 % identity to human chondroitin synthase 1 (hCSS1) and 18.0 % identity to mouse chondroitin synthase 1 (mCSS1), while the NGalNAc-T2 has 18.2 % to hCSS1 and 18.1% to mCSS1.

The mNGalNAc-T1 has 18.5 % identity to hCSS1 and 18.1 % identity to mCSS1, while the mNGalNAc-T2 has 18.1 % identity to hCSS1 and 18.8 % identity to mCSS1.

Therefore, it is recognized that the protein of the present invention is a novel one.

In addition, the protein of the present invention has the identity of 27 or more % to the amino acid sequence of SEQ ID NO: 1 or 3.

The protein of the present invention has the identity of 19 or more % to the amino acid sequence of SEQ ID NO: 26 or 28.

In addition, GENETYX is a genetic information processing software for nucleic acid analysis and protein analysis, which is capable of performing general homology analysis and multiple alignment analysis, as well as calculating a signal peptide, a site of promoter, and secondary structure. The program for homology analysis used herein adopts the Lipman-Pearson method (Lipman, D. J. & Pearson, W. R., Science, 277, 1435-1441 (1985)) which is frequently used as a high speed, highly sensitive method.

The amino acid sequences of the proteins and the DNA sequences encoding them disclosed herein can be wholly or

partially used to readily isolate genes encoding proteins having a similar physiological activity from that of other species using genetic engineering techniques including hybridization and nucleic acid amplification reactions such as PCR. In such cases, novel proteins encoded by these genes can also be included in the scope of the present invention.

Proteins of the present invention may contain an attached sugar chain if they have an amino acid sequence as defined above as well as the enzymatic activity described above.

More specifically, as described in Examples 2 and 5 below, from the search of an acceptor substrate to the protein of the present invention, said protein acts to transfer GalNAc to GlcNAc via a β1-4 linkage.

Furthermore specifically, the proteins of the present invention have at least one of the following properties (A)-(C), preferably all of these properties:

(A) Specificity of acceptor substrates

(a) When O-linked oligosaccharides are used as an acceptor substrate, said proteins have the activity of transferring N-acetylgalactosamine to GlcNAcβ1-6(Galβ1-3)GalNAcα-pNp (hereinafter, "core2-pNp"), GlcNAcβ1-3GalNAcα-pNp (hereinafter, "core3-pNp"), GlcNAcβ1-6GalNAcα-pNp (hereinafter, "core6-pNp") via a β1-4 linkage, wherein the abbreviations used are: GlcNAc, N-acetylglucosamine; GalNAc, N-acetylgalactosamine; Gal, galactose; pNp, p-nitrophenyl. Preferably, said proteins have the transferring activity to core6-pNp.

(b) When N-linked oligosaccharides are used as an

acceptor substrate, said proteins have the activity of transferring N-acetylgalactosamine to GlcNAc at the non-reducing end of said oligosaccharides via a β1-4 linkage, provided that said activity reduces when said oligosaccharides have the following properties:

(i) having fucose (Fuc) residues in the structure of said oligosaccharides; and

(ii) having one or more branched chains wherein GalNAc residues bind to GlcNAc residues at the non-reducing end.

(B) Optimum pH in enzymatic activity

The activity tends to be higher in pH 6.5 of MES (2-morpholineethanesulfonic acid) buffer. In HEPES ([4-(2-hydroxyethyl)-1-piperazinyl]ethanesulfonic acid) buffer, the activity tends to be higher in pH 6.75 for NGalNAc-T1 and pH 7.4 for NGalNAc-T2.

(C) Requirement of divalent ions

In NGalNAc-T1, the activity tends to be higher in the MES buffer including at least Mn^{2+}, or Co^{2+}, preferably Mn^{2+}. In NGalNAc-T2, the activity tends to be higher in the MES buffer including Mg^{2+}, Mn^{2+}, or Co^{2+}, preferably Mg^{2+}.

(2) Nucleic acids

Nucleic acids of the present invention include DNA in both single-stranded and double-stranded forms, as well as the RNA complements thereof. DNA includes, for example, native DNA, recombinant DNA, chemically synthesized DNA, DNA amplified by PCR and combinations thereof. The nucleic acid of the present invention is preferably a DNA.

The nucleic acids of the present invention are

nucleic acids (including the complement thereof) encoding the amino acids shown in SEQ ID NO: 1, 3, 26 or 28. Typically, the nucleic acids of the present invention have the nucleic acid sequence of SEQ ID NO: 2, 4, 27 or 29 (including the complements thereof), which are clones obtained in the working example below which shows simply an example of the present invention. It is well-known for a person skilled in the art that in native nucleic acids, there are minor mutants derived from various kinds of species which produce them and ecotypes and mutants from a presence of isozymes. Therefore, the nucleic acids of the present invention include, but are not limited to, the nucleic acids having the nucleic acid sequence shown in SEQ ID NO: 2, 4, 27 or 29. The nucleic acids of the present invention include all nucleic acids encoding the proteins of the present invention.

Particularly, the amino acid sequences of the proteins and the DNA sequences encoding them disclosed herein can be wholly or partially used to readily isolate nucleic acids encoding proteins having a similar physiological activity from that of other species using genetic engineering techniques including hybridization and nucleic acid amplification reactions such as PCR. In such cases, such nucleic acids can also be included in the scope of the present invention.

As used herein, "stringent conditions" means hybridization under conditions of moderate or high stringency. Specifically, conditions of moderate stringency can be readily determined by those having ordinary skill in

the art based on, for example, the length of the DNA. The basic conditions are shown by Sambrook et al., Molecular Cloning: A Laboratory Manual, 3rd edition, Vol. 1, 7.42-7.45 Cold Spring Harbor Laboratory Press, 2001 and include use of a prewashing solution for the nitrocellulose filters of 5 × SSC, 0.5 % SDS, 1.0 mM EDTA (pH 8.0), hybridization conditions of about 50 % formamide, 2 × SSC - 6 × SSC at about 40-50 °C (or other similar hybridization solution such as Stark's solution, in about 50 % formamide at about 42 °C), and washing conditions of 0.5 × SSC, 0.1 % SDS at about 60 °C. Conditions of high stringency can also be readily determined by those skilled in the art based on, for example, the length of the DNA. Generally, such conditions include hybridization and/or washing at a higher temperature and/or a lower salt concentration as compared with conditions of moderate stringency and are defined as hybridization conditions as above followed by washing in 0.2 × SSC, 0.1 % SDS at about 68 °C. Those skilled in the art will recognize that the temperature and the salt concentration of the washing solution can be adjusted as necessary according to factors such as the length of the probe.

Nucleic acid amplification reactions include reactions involving temperature cycles such as polymerase chain reaction (PCR) [Saiki R.K. et al., Science, 230, 1350-1354 (1985)], ligase chain reaction (LCR) [Wu D.Y. et al., Genomics, 4, 560-569 (1989); Barringer K.J. et al., Gene, 89, 117-122 (1990); Barany F., Proc. Natl. Acad. Sci. USA, 88, 189-193 (1991)] and transcription-based amplification [Kwoh

D.Y. et al., Proc. Natl. Acad. Sci. USA, 86, 1173-1177 (1989)] as well as isothermal reactions such as strand displacement amplification (SDA) [Walker G.T. et al., Proc. Natl. Acad. Sci. USA, 89, 392-396 (1992); Walker G.T. et al., Nuc. Acids Res., 20, 1691-1696 (1992)], self-sustained sequence replication (3SR) [Guatelli J.C., Proc. Natl. Acad. Sci. USA, 87, 1874-1878 (1990)], and Qß replicase system [Lizardi et al., BioTechnology, 6, 1197-1202 (1988)]. Other reactions such as nucleic acid sequence-based amplification (NASBA) using competitive amplification of a target nucleic acid and a variant sequence disclosed in European Patent No. 0525882 can also be used. PCR is preferred.

Homologous nucleic acids cloned by hybridization, nucleic acid amplification reactions or the like as described above have an identity of at least 50 % or more, preferably 60 % or more, more preferably 70 % or more, even more preferably 80 % or more, still more preferably 90 % or more, and most preferably 95 % or more to the nucleotide sequence of SEQ ID NO: 2, 4, 27 or 29 in the Sequence Listing.

The percentage identity of nucleic acid sequences may be determined by visual inspection and mathematical calculation. Alternatively, the percentage identity of two nucleic acid sequences can be determined by comparing sequence information using the GAP computer program, version 6.0 described by Devereux et al., Nucl. Acids Res., 12:387 (1984) which is available from the University of Wisconsin Genetics Computer Group (UWGCG). The preferred default parameters for the GAP program include: (1) a unary

comparison matrix (containing a value of 1 for identities and 0 for non-identities) for nucleotides, and the weighted comparison matrix of Gribskov and Burgess, Nucl. Acids Res., 14:6745 (1986), as described by Schwartz and Dayhoff, eds; Atlas of Protein Sequence and Structure, National Biomedical Research Foundation, pp. 353-358 (1979); (2) a penalty of 3.0 for each gap and an additional 0.10 penalty for each symbol in each gap; and (3) no penalty for end gaps. Other programs used by one skilled in the art of sequence comparison may also be used.

When the identity search of the nucleic acid of the present invention is performed using GENETYX (Genetyx Co.), the NGalNAc-T1 has 59.7 % identity to NGalNAc-T2, 81.4 % identity to mNGalNAc-T1, and 59.0 % identity to mNGalNAc-T2. The NGalNAc-T2 has 59.7 % identity to mNGalNAc-T1, and 83.4 % identity to mNGalNAc-T2. The mNGalNAc-T1 has 59.6 % identity to mNGalNAc-T2.

The NGalNAc-T1 has 44.6 % identity to hCSS1 and 46.0 % identity to mCSS1, while the NGalNAc-T2 has 47.3 % to hCSS1 and 47.9 % to mCSS1.

The mNGalNAc-T1 has 46.4 % identity to hCSS1 and 46.6 % identity to mCSS1, while mNGalNAc-T2 has 48.6 % identity to hCSS1 and 48.7 % identity to mCSS1.

Therefore, it is recognized that the nucleic acid of the present invention is a novel one.

In addition, the nucleic acid of the present invention has the identity of 48 or more % to the amino acid sequence of SEQ ID NO: 2 or 4.

The nucleic acid of the present invention has the

identity of 49 or more % to the amino acid sequence of SEQ ID NO: 27 or 29.

(3) Recombinant vectors and transformants

The present invention provides the recombinant vectors containing the nucleic acid of the present invention. Methods for integrating a DNA fragment of a nucleic acid of the present invention into a vector such as a plasmid are described in, for example, Sambrook, J. et al., Molecular Cloning, A Laboratory Manual (3rd edition), Cold Spring Harbor Laboratory, 1.1 (2001). Commercially available ligation kits (e.g., those available from Takara Shuzo Co., Ltd.) can be conveniently used. Thus obtained recombinant vectors (e.g., recombinant plasmids) are introduced into host cells (e.g., E. coli, TB1, LE392, or XL-1Blue, etc.).

Suitable methods for introducing a plasmid into a host cell include the use of calcium chloride or calcium chloride/rubidium chloride or calcium phosphate, electroporation, electro injection, chemical treatment with PEG or the like, and the use of a gene gun as described in Sambrook, J. et al., Molecular Cloning, A Laboratory Manual (3rd edition), Cold Spring Harbor Laboratory, 16.1 (2001).

Vectors can be conveniently prepared by linking a desired gene by a standard method to a recombination vector available in the art (e.g., plasmid DNA). Specific examples of suitable vectors include, but are not limited to, E. coli-derived plasmids such as pBluescript, pUC18, pUC19 and pBR 322.

In order to produce desired proteins, especially,

expression vectors are useful. The types of expression vectors are not specifically limited to those having the ability to express a desired gene in various prokaryotic and/or eukaryotic host cells to produce a desired protein, but preferably include expression vectors for E. coli such as pQE-30, pQE-60, pMAL-C2, pMAL-p2, pSE420; expression vectors for yeasts such as pYES2 (genus Saccharomyces), pIC3.5K, pPIC9K, pA0815 (all belonging to genus Pichia); and expression vectors for insects such as pBacPAK8/9, pBK283, pVL1392, pBlueBac4.5.

A transformant can be produced by introducing a desired expression vector into a host cell. The host cells employed are not specifically limited to those having the ability to be compatible to the expression vector of the present invention and to be able to be transformed, but various kinds of cells such as native cells are usually used in the art or recombinant cells are artificially established. For example, bacteria (genus Escherichia, genus Bacillus), yeasts (genus Saccharomyces, genus Pichia, etc.), mammalian cells, insect cells, and plant cells are exemplified.

The host cells are preferably E. coli, yeasts and insect cells, which are exemplified as E. coli (M15, JM109, BL21, etc.), yeasts (INVSc1 (genus Saccharomyces), GS115, KM71 (genus Pichia), etc.), and insect cells (BmN4, bombic larva, etc.). Examples of animal cells are mouse, Xenopus, rat, hamster, monkey or human derived cells or culture cell lines established from these cells. More specifically, the host cell is preferably COS cell which is a cell line derived from a kidney of monkey.

When a bacterium, especially E. coli is used as a host cell, the expression vector typically consists of at least a promoter/operator region, a start codon, a gene encoding a desired protein, a stop codon, a terminator and a replicable unit.

When a yeast, plant cell, animal cell or insect cell is used as a host cell, the expression vector typically preferably contains at least a promoter, a start codon, a gene encoding a desired protein, a stop codon and a terminator. It may also contain a DNA encoding a signal peptide, an enhancer sequence, untranslated regions at the 5' and 3' ends of a desired gene, a selectable marker region or a replicable unit, etc., if desired.

Preferred start codons in vectors of the present invention include a methionine codon (ATG). Stop codons include commonly used stop codons (e.g., TAG, TGA, TAA).

The replicable unit means DNA capable of replicating the entire DNA sequence in a host cell, such as natural plasmids, artificially modified plasmids (plasmids prepared from natural plasmids), synthetic plasmids, etc. Preferred plasmids include plasmid pQE30, pET or pCAL or their artificial variants (DNA fragments obtained by treating pQE30, pET or pCAL with suitable restriction endonucleases) for E. coli; plasmid pYES2 or pPIC9K for yeasts; and plasmid pBacPAK8/9 for insect cells.

Enhancer sequences and terminator sequences may be those commonly used by those skilled in the art such as those derived from SV40.

As for selectable markers, those commonly used can be

used by standard methods. Examples are genes resistant to antibiotics such as tetracycline, ampicillin, kanamycin, neomycin, hygromycin or spectinomycin.

Expression vectors can be prepared by linking at least a promoter, a start codon, a gene encoding a desired protein, a stop codon and a terminator region as described above to a suitable replicable unit in series into a circle. While carrying out the linking process, a suitable DNA fragment (such as a linker or another restriction site) can be used by standard methods such as digestion with a restriction endonuclease or ligation with T4 DNA ligase, if desired.

Introduction [transformation (transduction)] of expression vectors of the present invention into host cells can be performed by using known techniques.

For example, bacteria (such as E. coli, Bacillus subtilis) can be transformed by the method of Cohen et al. [Proc. Natl. Acad. Sci. USA, 69, 2110 (1972)], the protoplast method [Mol. Gen. Genet., 168, 111 (1979)] or the competent method [J. Mol. Biol., 56, 209 (1971)]; Saccharomyces cerevisiae can be transformed by the method of Hinnen et al [Proc. Natl. Acad. Sci. USA, 75, 1927 (1978)] or the lithium method [J.B. Bacteriol., 153, 163 (1983)]; plant cells can be transformed by the leaf disc method [Science, 227, 129 (1985)] or electroporation [Nature, 319, 791 (1986)]; animal cells can be transformed by the method of Graham [Virology, 52, 456 (1973)]; and insect cells can be transformed by the method of Summers et al. [Mol. Cell. Biol., 3, 2156-2165 (1983)].

(4) Isolation/purification of proteins

Proteins of the present invention can be expressed (produced) by culturing transformed cells containing an expression vector prepared as described above in a nutrient medium. The nutrient medium preferably contains a carbon, inorganic nitrogen or organic nitrogen source necessary for the growth of host cells (transformants). Examples of carbon sources include glucose, dextran, soluble starch, sucrose and methanol. Examples of inorganic or organic nitrogen sources include ammonium salts, nitrates, amino acids, corn steep liquor, peptone, casein, beef extract, soybean meal and potato extract. If desired, other nutrients (e.g., inorganic salts such as sodium chloride, calcium chloride, sodium dihydrogen phosphate and magnesium chloride; vitamins; antibiotics such as tetracycline, neomycin, ampicillin and kanamycin) may be contained. Incubation of cultures takes place by techniques known in the art. Culture conditions such as temperature, the pH of the medium and the incubation period are appropriately selected to produce a protein of the present invention in mass.

Proteins of the present invention can be obtained from the resulting cultures as follows. That is, when proteins of the present invention accumulate in host cells, the host cells are collected by centrifugation or filtration or the like and suspended in a suitable buffer (e.g., a buffer such as a Tris buffer, a phosphate buffer, an HEPES buffer or an MES buffer at a concentration of about 10 M -

100 mM desirably at a pH in the range of 5.0 - 9.0, though the pH depends on the buffer used), then the cells are disrupted by a method suitable for the host cells used and centrifuged to collect the contents of the host cells. When proteins of the present invention are secreted from host cells, the host cells and culture medium are separated by centrifugation or filtration or the like to give a culture filtrate. The disruption solution of the host cells or the culture filtrate can be used to isolate/purify a protein of the present invention directly or after ammonium sulfate precipitation and dialysis. An isolation/purification method is as follows. When the protein of interest is tagged with 6 x histidine, GST, maltose-binding protein or the like, conventional methods based on affinity chromatography suitable for each tag can be used. When the protein of the present invention is produced without using these tags, the method described in detail in the examples below based on ion exchange chromatography can be used, for example. These methods may be combined with gel filtration chromatography, hydrophobic chromatography, isoelectric chromatography or the like.

N-acetylgalactosamine is transferred by the action of proteins of the present invention on glycoprotein, oligosaccharide, polysaccharide or the like having N-acetylglucosamine. Thus, proteins of the present invention can be used to modify a sugar chain of a glycoprotein or to synthesize a sugar. Moreover, the proteins can be administered as immunogens to an animal to prepare antibodies against said proteins, and said antibodies can be

used to determine said proteins by immunoassays. Thus, proteins of the present invention and the nucleic acids encoding them are useful in the preparation of such immunogens.

Further, proteins of the present invention can comprise peptides added to facilitate purification and identification. Such peptides include, for example, poly-His or the antigenic identification peptides described in US Patent No. 5,011,912 and in Hopp et al., Bio/Technology, 6:1204, 1988. One such peptide is the FLAG® peptide, Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys (SEQ ID NO: 30) which is highly antigenic and provides an epitope reversibly bound by a specific monoclonal antibody, enabling rapid assay and facile purification of expressed recombinant protein. A murine hybridoma designated 4E11 produces a monoclonal antibody that binds the FLAG® peptide in the presence of certain divalent metal cations, as described in US Patent No. 5,011,912 hereby incorporated by reference. The 4E11 hybridoma cell line has been deposited with the American Type Culture Collection under Accession No. HB 9259. Monoclonal antibodies that bind the FLAG® peptide are available from Eastman Kodak Co., Scientific Imaging Systems Division, New Haven, Connecticut.

Specifically, the cDNA of the FLAG is inserted into an expression vector expressing a protein of the present invention to express the FLAG-tagged protein, after which the expression of the protein of the present invention can be confirmed by an anti-FLAG antibody.

(5) Analytical nucleic acid

According to the present invention, a nucleic acid which hybridizes to the nucleic acids of the present invention (hereinafter referred to as "analytical nucleic acid") is provided. The analytical nucleic acid of the present invention includes, but is not limited to, typically, native or synthesized fragments derived from nucleic acid encoding the protein of the present invention. As used herein, the term "analytical" includes any of detection, amplification, quantitative and semi-quantitative assays.

(a) Primers

When analytical nucleic acids of the present invention are used as primers for nucleic acid amplification reactions, the analytical nucleic acids of the present invention are oligonucleotides prepared by a process comprising:

selecting two regions from the nucleotide sequence of a gene encoding a protein of SEQ ID NO: 1, 3, 26 or 28 to satisfy the conditions that:

1) each region should have a length of 15-50 bases; and

2) the proportion of G + C in each region should be 40-70 %;

generating a single-stranded DNA having a nucleotide sequence identical to or complementary to that of said region or generating a mixture of single-stranded DNAs taking into account degeneracy of the genetic code so that the amino acid residue encoded by said single-stranded DNA

is retained, and, as necessary, generating the single-stranded DNA containing a modification without affecting the binding specificity to the nucleotide sequence of the gene encoding said protein.

Primers of the present invention preferably have a sequence homologous to that of a partial region of a nucleic acid of the present invention, but one to two bases may be mismatched.

Primers of the present invention contain 15 bases or more, preferably 18 bases or more, more preferably 21 bases or more, and 50 bases or fewer bases.

The primer of the present invention has typically the nucleic acid sequence selected of a group consisting of SEQ ID NO: 20, 21, 23 and 24, and can be used as a single primer or a suitably combined pair of primers. These nucleotide sequences were designed based on amino acid sequence of SEQ ID 1 or 3 as a PCR primer for cloning gene fragments encoding each protein. The sequence is a primer mixed with all nucleic acids capable of encoding said amino acids.

(b) Probes

When analytical nucleic acids of the present invention are used as probes, the analytical nucleic acids of the present invention preferably have a sequence homologous to that of a total or partial region of the nucleotide sequence of SEQ ID NO: 2, 4, 27 or 29, and further, may have a mismatch of one or two bases. The probes of the present invention have a length of 15 bases and more, preferably 20 bases and more, and within a full

length of the encoding region, that is, 3120 bases (corresponding to SEQ ID NO: 2), 2997 bases (corresponding to SEQ ID NO: 4), 3105 bases (corresponding to SEQ ID NO: 27), or 2961 bases (corresponding to SEQ ID NO: 29). The probes have typically the nucleic acid sequence shown in SEQ ID NO: 22 or 25. The probes may be obtained from native nucleic acid treated with restriction enzymes, or may be synthesized oligonucleotides.

Probes of the present invention include labeled probes having a label such as a fluorescent, radioactive or biotinylation label to detect or confirm that the probes have hybridized to a target sequence. The presence of a nucleic acid to be tested in an analyte can be determined by immobilizing the nucleic acid to be tested or an amplification product thereof, hybridizing it to a labeled probe, and after washing, measuring the label bound to the solid phase. Alternatively, it can also be determined by immobilizing the analytical nucleic acid, hybridizing to the nucleic acid to be tested and detecting the nucleic acid to be tested coupled to the solid phase with a labeled probe or the like. In the latter case, the immobilized analytical nucleic is also referred to as a probe.

Generally, nucleic acid amplification methods such as PCR can be readily performed because they are per se well known in the art, and reagent kits and apparatus for them are also commercially available. When a nucleic acid amplification method is performed using a pair of analytical nucleic acids of the present invention described above as primers and a nucleic acid to be tested as the template, the

presence of the nucleic acid to be tested in a sample can be known by detecting an amplification product because the nucleic acid to be tested is amplified while no amplification occurs when the nucleic acid to be tested is not contained in the sample. The amplification product can be detected by electrophoresing the reaction solution after amplification, staining the bands with ethidium bromide, immobilizing the amplification product after electrophoresis to a solid phase such as a nylon membrane, hybridizing the immobilized product with a labeled probe that specifically hybridizes to the nucleic acid to be tested, and washing the hybridization product and then detecting said label. Further, the amount of the nucleic acid to be tested in a sample can also be determined by the so-called real-time PCR detection using a quencher fluorescent dye and a reporter fluorescent dye. This method can also be readily carried out using a commercially available real-time PCR detection kit. The nucleic acid to be tested can also be semi-quantitatively assayed based on the intensity of electrophoretic bands. The nucleic acid to be tested may be mRNA or cDNA reversely transcribed from mRNA. When mRNA is to be amplified as the nucleic acid to be tested, the NASBA methods (3SR, TMA) can also be adopted using said pair of primers. The NASBA methods can be readily performed because they are per se well known and kits for them are commercially available.

(c) Microarrays

Analytical nucleic acids of the present invention can

be used as microarrays. Microarrays are means for enabling rapid large-scale data analysis of genomic functions. Specifically, a labeled nucleic acid is hybridized to a number of different nucleic acid probes immobilized in high density on a solid substrate such as a glass substrate, a signal from each probe is detected and the collected data are analyzed. As used herein, the "microarray" means an array of an analytical nucleic acid of the present invention on a solid substrate such as a membrane, filter, chip or glass surface.

(6) Antibodies

An antibody that is immunoreactive with the protein of the present invention is provided herein. Such an antibody specifically binds to the polypeptide via the antigen-binding site of the antibody (as opposed to non-specific binding). Therefore, as set forth above, proteins of SEQ ID NOs: 1 and 3, fragments, variants, and fusion proteins and the like can be used as "immunogens" in producing antibodies immunoreactive therewith. More specifically, the proteins, fragments, variants, and fusion proteins and the like include the antigenic determinants or epitopes to induce the formation of an antibody. Such antigenic determinants or epitopes may be either linear or conformational (discontinuous). In addition, said antigenic determinants or epitopes may be identified by any methods known in the art.

Therefore, one aspect of the present invention relates to the antigenic epitopes of the protein of the

present invention. Such epitopes are useful raising antibodies, in particular monoclonal antibodies, as described in more detailed below. Additionally, epitopes from the protein of the present invention can be used as research reagents, in assays, to purify specific binding antibodies from substances such as polyclonal sera or supernatants from cultured hybridomas. Such epitopes or variants thereof can be produced using techniques known in the art such as solid-phase synthesis, chemical or enzymatic cleavage of a protein, or by using recombinant DNA technology.

As for antibodies which can be induced by the proteins of the present invention, both polyclonal and monoclonal antibodies can be prepared by conventional techniques, whether a whole body or a part of said proteins have been isolated, or the epitopes have been isolated. See, for example, Monoclonal Antibodies, Hybridomas: A New Dimension in Biological Analyses, Plenum Press, NY, 1980.

Hybridoma cell lines that produce monoclonal antibodies specific for the proteins of the present invention are also contemplated herein. Such hybridomas can be produced and identified by conventional techniques. One method for producing such a hybridoma cell line comprises immunizing an animal with a protein of the present invention; harvesting spleen cells from the immunized animal; fusing said spleen cells to a myeloma cell line, thereby generating hybridoma cells; and identifying a hybridoma cell line that produces a monoclonal antibody that binds said protein. The monoclonal antibodies can be

recovered by conventional techniques.

The antibodies of the present invention include chimeric antibodies such as humanized versions of murine monoclonal antibodies. Such humanized antibodies can be prepared by known techniques and offer the advantages of reduced immunogenicity when the antibodies are administered to humans. In one embodiment, a humanized monoclonal antibody comprises the variable region of a murine antibody (or just the antigen-binding site thereof) and a constant region derived from a human antibody. Alternatively, a humanized antibody fragment can comprise the antigen-binding site of a murine monoclonal antibody and a variable region fragment (lacking the antigen-biding site) derived from a human antibody.

The present invention includes antigen-binding antibody fragments that can be also generated by conventional techniques. Such fragments include, but are not limited to, Fab and $F(ab')_2$ as an example. Antibody fragments generated by genetic engineering techniques and derivatives thereof are also provided.

In one embodiment, the antibody is specific to the protein of the present invention, and it does not cross-react with other proteins. Screening procedures by which such antibodies can be identified are publicly known, and may involve, for example, immunoaffinity chromatography.

The antibodies of the invention can be used in assays to detect the presence of the protein or fragments of the present invention, either in vitro or in vivo. The antibodies also can be used in purifying proteins or

fragments of the present invention by immunoaffinity chromatography.

Further, a binding partner such as an antibody that can block binding of a protein of the present invention to an acceptor substrate can be used to inhibit a biological activity rising from such a binding. Such a blocking antibody may be identified by any suitable assay procedure, such as by testing the antibody for the ability to inhibit binding of said protein to specific cells expressing the acceptor substrate. Alternatively, a blocking antibody can be identified in assays for the ability to inhibit a biological effect that results from a protein of the present invention binding to the binding partner of target cells.

Such an antibody can be used in an in vitro procedure, or administered in vivo to inhibit a biological activity mediated by the entity that generated the antibody. Disorders caused or exacerbated (directly or indirectly) by the interaction of a protein of the present invention with a binding partner thus can be treated. A therapeutic method involves in vivo administration of a blocking antibody to a mammal in an amount effective to inhibit a binding partner-mediated biological activity. Monoclonal antibodies are generally preferred for use in such therapeutic methods. In one embodiment, an antigen-binding antibody fragment is used.

(7) Cancer markers and methods for detection

The protein or nucleic acids of the present invention can be used as a cancer marker, and be applied to diagnosis and treatment of cancers and the like. As used herein, the

term "cancer" means typically all malignant tumors, and includes disease conditions with said malignant tumors. "Cancer" includes, but is not limited to, lung cancer, liver cancer, kidney cancer and leukemia.

"Cancer marker" used herein means the protein and nucleic acids of the present invention that express more than those of a non-cancerous biological sample, when a biological sample is cancerous. In addition, "biological sample" includes tissues, organs, and cells. Blood is preferable, pathological tissue is more preferable.

Specifically, when the protein of the present invention is used as a cancer marker, a method for detection of the present invention includes the steps: (a) quantifying said protein in a biological sample; and (b) estimating that the biological sample is cancerous in the case that the quantity value of said protein in the biological sample is more than that in a control biological sample. In said method for detection, the antibody of the present invention can be used to quantify said protein of the biological sample. According to the present invention, generally, the method for qualifying the protein is not limited to the above methods and can use quantity methods know in the art such as ELISA, Western Blotting. A ratio of the quantity value is preferably 1.5 times or more, more preferably 3 times or more, and even more preferably 10 times or more.

On the other hand, when the nucleic acid of the present invention is used as a cancer marker, a method for detection of the present invention includes the steps of: (a) quantifying said nucleic acid in a biological sample;

and (b) estimating that the biological sample is cancerous in the case that the quantity value of said nucleic acid in the biological sample is 1.5 times or more than that of a control biological sample. Preferably, the steps comprise (a) hybridizing at least one of said analytical nucleic acids to said nucleic acid in the biological sample; (b) amplifying said nucleic acid; (c) hybridizing said nucleic acids to the amplification product; (d) quantifying a signal rising from said amplification product and said analytical nucleic acid hybridized; and (e) estimating that the biological sample is cancerous in the case that the quantity value of said signal is 1.5 times or more than that of a corresponding signal of a control biological sample.

More specifically, as described in the example below, canceration can be estimated by determination of a ratio of expression level of the nucleic acids in cancerous tissue and normal tissue by quantitative PCR. According to the present invention, the quantification of the nucleic acid is not limited to this, and for example, RT-PCR, northern blotting, dot blotting or DNA microarray may be used. In such quantification, nucleic acids of genes present generally and broadly in same tissue and the like such as nucleic acids encoding glyceraldehyde-3-phosphate dehydrogenase (GAPDH), β-actin are used as a control. A quantity ratio to be estimated as canceration is preferably 1.5 or more, more preferably 3 or more, even more preferably 10 or more.

The following examples further illustrate the present

invention without, however, limiting the invention thereto.

Examples

Example 1 Preparation of the human protein of the present invention

1. Search through a genetic database and determination of the nucleic acid sequence of a novel N-acetylgalactosamine transferase

A search of similar genes through a genetic database was performed by use of the genes for existing β-1,4-galactose transferases. The sequences used were SEQ ID NOs: AL161445, AF038660, AF038661, AF022367, AF038663, AF038664 in the genes for β-1,4-galactose transferases. The search was performed using a program such as Blast [Altschul et al., J. Mol. Biol., 215, 403-410 (1990)].

As a result, GenBank Accession No. N48738 was found as an EST sequence, and GenBank Accession No. AC006205 was found as a genome sequence. As a further result, it is considered that both sequences comprise disparate genes (hereinafter, the genes comprising N48738 and AC006205 refer to NGalNAc-T1 and NGalNAc-T2, respectively). Since the translation initiation sites of both genes were unknown, it was impossible to predict the full length of the genes. Marathon-Ready cDNA (Human Brain or Stomach) from CLONTECH was used for obtaining the information of coding regions (5' RACE: Rapid Amplification of cDNA Ends) and cloning.

Obtaining information of coding region of NGalNAc-T1

AP1 primer included in Marathon cDNA (a DNA fragment having adaptors AP1 and AP2 at both ends) and primer K12R6 generated within the identified sequence part (5'-GCT CCT GCA GCT CCA GCT CCA-3') (SEQ ID NO: 5) were used for PCR (30 cycles of 94 °C for 20 seconds, 60 °C for 30 seconds and 72 °C for 2 minutes). Further, AP2 primer included in Marathon cDNA and primer K12R5 generated within the identified sequence part (5'-AAG CGA CTC CCT CGC GCC GAG T-3') (SEQ ID NO: 6) were used for nested PCR (30 cycles of 94 °C for 20 seconds, 60 °C for 30 seconds and 72 °C for 2 minutes). A fragment of about 0.6 kb obtained as a result was purified by a common method, and the nucleic acid sequence was analyzed. However, since a transmembrane sequence special to glycosyl transferases (hydrophobic 20 amino acids) could have appeared, an EST sequence (GenBank Accession No. PF058197) was discovered based on the obtained sequence and the nucleic acid sequence of NGalNAc-T2 described later by search through genome database. Based on the information of nucleic acid sequence, RT-PCR was performed using two primers (K12F101: 5'-ATG CCG CGG CTC CCG GTG AAG AAG-3' (SEQ ID NO: 7) and K12R5) and the amplification was confirmed. Therefore, it was explained that this EST sequence and the sequence obtained by 5' RACE exist on one mRNA. The full length of nucleotide sequence (3120 bp) was shown in SEQ ID NO: 2.

Obtaining information of coding region of NGalNAc-T2

AP1 primer included in Marathon cDNA (a DNA fragment

having adaptors AP1 and AP2 at both ends) and primer K13-R3 generated within the identified sequence part (5'-CAA CAG TTC AAG CTC CAG GAG GTA-3' (SEQ ID NO: 8)) were used for PCR (30 cycles of 94 °C for 20 seconds, 60 °C for 30 seconds and 72 °C for 2 minutes). Further, AP2 primer included in Marathon cDNA and primer K13R2 generated within the identified sequence part (5'-CTG ACG CTT TTC CAC GTT CAC AAT-3'(SEQ ID NO: 9)) were used for nested PCR (30 cycles of 94 °C for 20 seconds, 60 °C for 30 seconds and 72 °C for 2 minutes). A fragment of about 1.0 kb obtained as a result was purified by a common method, and the nucleic acid sequence was analyzed. Further, a coding region of a protein was determined. However, since a transmembrane sequence special to glycosyl transferases (hydrophobic 20 amino acids) could have appeared, further 3 times 5' RACE was performed. The primers used here are shown in Table 2.

As a result, the obtained full length of nucleotide sequence (2997 bp) was shown in SEQ ID NO: 4.

Table 2 Various primers used in RACE

Second 5' RACE primers

K13 R6 5'-CAC CCC GTC TCT GCT CTG CGA T-3'(SEQ ID NO: 10)

K13 R5 5'-GTC TTC CTG GGG CTG TCA CCA-3' (SEQ ID NO: 11)

Third 5' RACE primers

K13 R7 5'-CAC CTC ATC CAT CTG TAG GAA CGT-3'(SEQ ID NO: 12)

K13 R8 5'-CTG TCG CCA TGC AAC TTC CAC GT-3' (SEQ ID NO: 13)

Fourth 5' RACE primers

K13 R12 5'-AAT GTC GTG GTC CTC GAG GCT CA-3' (SEQ ID NO: 14)

K13 R11 5'-GAT GGT AGA ACT GGA GGT GTG GAT-3'(SEQ ID NO: 15)

2. Integration of GalNAc-T gene into an expression vector

To prepare an expression system of GalNAc-T, a portion of GalNAc-T gene was first integrated into pFLAG-CMV1 (Sigma).

Integration of NGalNAc-T1 into pFLAG-CMV1

A region corresponding to amino acids 62-1039 of SEQ ID NO: 1 or 2 was amplified by LA Taq DNA polymerase (Takara Shuzo) using Marathon cDNA (Human Brain) as a template, forward primer K12-Hin-F2: 5'-CCC AAG CTT CGG GGG GTC CAC GCT GCG CCA T-3' (SEQ ID NO: 16), and reverse primer K12-Xba-R1: 5'-GCT CTA GAC TCA AGA CGC CCC CGT GCG AGA-3' (SEQ ID NO: 17). The fragment was digested at restriction sites (HindIII and XbaI) included in the primers, and inserted into pFLAG-CMV1 digested with Hind III and XbaI by use of Ligation High (Toyobo) to prepare pFLAG-NGalNAc-T1.

Integration of NGalNAc-T2 into pFLAG-CMV1

A region corresponding to amino acids 57-998 of SEQ ID NO: 3 or 4 was amplified by LA Taq DNA polymerase (Takara Shuzo) using Marathon cDNA (Human Stomach) as a template, forward primer K13-Eco-F1: 5'-GGA ATT CGA GGT ACG GCA GCT GGA GAG AA-3' (SEQ ID NO: 18), and reverse primer K13-Sal-R1: 5'-ACG CGT CGA CCT ACA GCG TCT TCA TCT GGC GA-3' (SEQ ID NO: 19). This fragment was digested at restriction sites (EcoRI and SalI) included in the primers, and inserted temporally into pcDNA3.1 digested with EcoRI and SalI. This

was digested with EcoRI and PmeI. The fragment including the active site of NGalNAc-T2 was inserted at the EcoRI-EcoRV site ofpFLAG-CMV1 using Ligation High (Toyobo Co.) to prepare pFLAG-NGalNAc-T2.

3. Transfection and expression of recombinant enzymes

15 μg of pFLAG-NGalNAc-T1 or pFLAG-NGalNAc-T2 was induced into 2 X 10^6 of COS-1 cells which were cultured overnight in DMEM (Dulbecco's modified Eagle's medium) including 10 % FCS (fetal calf serum), using Lipofectamine 2000 (Invitrogen Co.) as a protocol provided by the same company. A supernatant of 48-72 hours was collected. The supernatant was mixed with NaN_3 (0.05 %), NaCl (150 mM), $CaCl_2$ (2 mM) and an anti-M1 resin (Sigma Co.) (50 μl), and the mixture was stirred overnight at 4 °C. The solution of reaction mixture was centrifuged (3000 rpm, 5 min, 4 °C) to collect a pellet. The pellet was combined with 900 μl of 2 mM $CaCl_2$/TBS and re-centrifuged (2000 rpm, 5 min, 4 °C), after which the pellet was suspended in 200 μl of 1 mM $CaCl_2$/TBS to give a sample for assaying activity (NGalNAc-T1 or NGalNAc-T2 enzyme solution).

The enzyme was subjected to conventional SDS-PAGE and Western blotting, and the expression of the intended protein was confirmed. Anti FLAG M2-peroxydase (A-8592, SIGMA Co.) was used as an antibody.

Example 2 Assay of activity using the enzyme of the present invention

1. Search for donor substrates

A search for a donor substrate of the enzyme of the present invention was performed on various mono-saccharide acceptor substrates, using 5 ml of enzyme solution and various acceptor substrates.

The acceptor substrates were prepared so that each of Gal-α-pNp, Gal-β-oNp, GalNAc-α-Bz, GalNAc-β-pNp, GlcNAc-α-pNp, GlcNAc-β-pNp, Glc-α-pNp, Glc-β-pNp, GlcA-β-pNp, Fuc-α-pNp, Man-α-pNp (thereinbefore, CALBIOCHEM Co.), Xyl-α-pNp, Xyl-β-pNp (thereinbefore, SIGMA Co.) was included in 2.5 nmol/20 μl. Further, the solutions of various donor substrates (UDP-GalNAc, UDP-GlcNAc, UDP-Gal, GDP-Man, UDP-GlcA, UDP-Xyl and GDP-Fuc, thereinbefore, SIGMA Co.) are shown in Table 3.

Table 3

GalNAc-T	
MES or HEPES (pH 5.5 -	50 mM
UDP-GalNAc	0.5 mM
UDP-[14C]GalNAc	2 nCi/ul
MnCl2	20 mM
Triron X-100	0.5%

GlcNAc-T	
HEPES (pH 7.0 or 7.5)	14 mM
UDP-GlcNAc	0.5 mM
UDP-[14C]GlcNAc	2 nCi/ul
MnCl2	10 mM
Triron CF-54	0.5%
ATP	0.75 mM

Gal-T	
HEPES (pH 7.0 or 7.5)	14 mM
UDP-Gal	0.25 mM
UDP-[14C]Gal	2.5 nCi/ul
MnCl2	10 mM
ATP	0.75 mM

GlcA-T	
MES (pH 7.0)	50 mM
UDP-GlcA	0.25 mM
UDP-[14C]GlcA	2 nCi/ul
MnCl2	10 mM

Xyl-T	
MES (pH 7.0)	50 mM
UDP-Xyl	0.25 mM
UDP-[14C]Xyl	1 nCi/ul
MnCl2	10 mM

Fuc-T	
cacodylate buffer (pH 7.0	50 mM
GDP-[14C]Fuc	1 nCi/ul
MnCl2	10 mM
ATP	5 mM

Man-T	
Tris (pH 7.2)	50 mM
GDP-[14C]Man	2 nCi/ul
MnCl2	10 mM
Triton X-100	0.6%

All of reaction times were 16 hours. After reaction, non-reactive acceptor substrates with radioactivity were removed with SepPack C18 column (Waters CO.), and radioactivity from donor substrates integrated into acceptor substrates was determined with a liquid scintillation counter. Consequently, there appeared little background even in UDP-GlcA using each of NGalNAc-T1 and NGalNAc-T2, however, the highest activity was detected in the case of UDP-GalNAc as a donor substrate.

2. Search for acceptor substrates

Further, in order to investigate acceptors, reactions were performed using each acceptor (10 nmol/20 μl) by itself. As a result, significant radioactivity was detected in the case of GlcNAc-β-pNp (NGalNAc-T1: 256.26 dpm, NGalNAc-T2: 1221.22 dpm). Based on the above results, it was explained that both of NGalNAc-T1 and NGalNAc-T2 are glycosyl transferases capable of transferring GalNAc to GlcNAc-T.

3. Study of optimum pH

As described above, it was explained that NGalNAc-T1 and NGalNAc-T2 are glycosyl transferases which transfer GalNAc to GlcNAc. Thereat, the optimum pH of both enzymes was studied. The buffer solutions used are MES (pH 5.5, 6.0, 6.26, 6.5, 6.75), HEPES (pH 6.75, 7.0, 7.4). As a result, as shown in Table 4, the activity tends to be higher in pH 6.5 of MES buffer for both NGalNAc-T1 and NGalNAc-T2.

Table 4 A result of optimum pH in enzymatic activity of NGalNAc-T1 and NGalNAc-T2

NGalNAc-T1

pH	Incorporation of radioactivity (A)	Blank (B)	(A) - (B)
MES buffer (pH 5.5)	339.76	263.21	76.55
MES buffer (pH 6.0)	321.04	263.21	57.83
MES buffer (pH 6.26)	636.34	263.21	373.13
MES buffer (pH 6.5)	1767.72	263.21	1504.51
MES buffer (pH 6.75)	923.92	263.21	660.71
HEPES buffer (pH 6.75)	1685.06	263.21	1421.85
HEPES buffer (pH 7.0)	1138.38	263.21	875.17
HEPES buffer (pH 7.4)	2587.48	263.21	2324.27

(dpm)

NGalNAc-T2

pH	Incorporation of radioactivity (A)	Blank (B)	(A) - (B)
MES buffer (pH 5.5)	336.20	263.21	72.99
MES buffer (pH 6.0)	341.92	263.21	78.71
MES buffer (pH 6.26)	339.50	263.21	76.29
MES buffer (pH 6.5)	753.62	263.21	490.05
MES buffer (pH 6.75)	529.24	263.21	266.03
HEPES buffer (pH 6.75)	915.16	263.21	651.95
HEPES buffer (pH 7.0)	786.70	263.21	523.49
HEPES buffer (pH 7.4)	586.32	263.21	323.11

(dpm)

In addition, the value (263.21 dpm) of MES (pH 6.75) was adopted as a blank value in the case of a non-enzyme. Further, when pH of HEPES buffer was 7.4 for NGalNAc-T1 and 6.75 for NGalNAc-T2, the highest value was shown. However, the activity did not always increase even when pH increase.

Hereinafter, MES (pH 6.5) was used in each of experiments.

4. Studying requirements of divalent cations

Generally, glycosyl transferases require frequently divalent cations. The activity of each enzyme was studied by adding various divalent cations. Consequently, the high values were represented when Mn^{2+} in NGalNAc-T1, and Mg^{2+}, Mn^{2+} and Co^{2+} in NGalNAc-T2 were added (see Table 5). Regarding this, both enzymes showed the activity due to adding EDTA which is a chelating agent. From the above results, it was explained that both enzymes require divalent cations.

Table 5 A result of requirements of divalent cations in the activity of NGalNAc-T1 and NGalNAc-T2

NGalNAc-T1

Divalent cations etc.	Incorporation of radioactivity (A)	Blank (B)	(A) - (B)
$MnCl_2$	519.47	263.21	256.26
$MgCl_2$	256.36	263.21	-6.85
$ZnCl_2$	210.29	263.21	-52.92
$CaCl_2$	230.78	263.21	-32.43
$CuCl_2$	278.77	263.21	15.56
$CoCl_2$	240.91	263.21	-22.30
$CdSO_4$	203.39	263.21	-59.82
EDTA	242.38	263.21	-20.83

(dpm)

NGalNAc-T2

pH	Incorporation of radioactivity (A)	Blank (B)	(A) - (B)
$MnCl_2$	1484.43	263.21	1221.22
$MgCl_2$	3124.16	263.21	2860.95
$ZnCl_2$	187.59	263.21	-75.62
$CaCl_2$	217.83	263.21	-45.38
$CuCl_2$	218.35	263.21	-44.86
$CoCl_2$	1130.63	263.21	867.42
$CdSO_4$	217.92	263.21	-45.29
EDTA	235.28	263.21	-27.93

(dpm)

Example 3 Expression analysis in various human tissues

The expression levels of said gene was quantified by quantitative PCR using cDNA of normal human tissues. The cDNA of normal tissues which was reversely transcribed from total RNA (CLONETECH Co.) was used. As for cell lines, total RNA therefrom was extracted, and cDNA was prepared by conventional methods and was used. The quantitative expression analysis of NGalNAc-T1 was performed using primers: K12-F3 (5'-ctg gtg gat ttc gag agc ga-3' (SEQ ID NO: 20)) and K12-R3 (5'-tgc cgt cca gga tgt tgg-3' (SEQ ID NO: 21)), and probe: K12-MGB3 (5'-gcg gta gag gac gcc-3' (SEQ ID NO: 22)). The quantitative expression analysis of NGalNAc-T2 was performed using primers: K13-F3 (5'-atc gtc atc act gac tat agc agt ga-3' (SEQ ID NO: 23)) and K13-R3 (5'-gaa tgg cat cga tga ctc cag-3' (SEQ ID NO: 24)), and probe: K13-MGB3 (5'-ctc gtg aag gac ccg ca-3' (SEQ ID NO:

25)). A prove with a minor groove binder (Applied Biosystems Co.) was used. Universal PCR Master Mix was used as enzyme and reaction solution, and 25 ml of the reaction solution was quantified with ABI PRISM 7700 Sequence Detection System (together, Applied Biosystems Co.). Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was used as a standard gene for quantification. A calibration curve for quantification was made by using a template DNA at a known concentration, and the expression level of said gene was normalized. Further, pFLAG-NGalNAc-T1 and pFLAG-NGalNAc-T2 were used as standard DNAs of NGalNAc-T1 and NGalNAc-T2. The reaction temperature was 50 °C for 2 min, 95 °C for 10 min, followed by 50 cycles of 95 °C for 15 sec, 60 °C for 1 min. The result is shown in Figure 1. It was explained that the amounts of expressions of NGalNAc-T1 and NGalNAc-T2 were high in the nervous system, stomach and spermary, respectively.

Example 4 Expression analysis of human cancerous tissue

The expression levels of both genes of human lung cancerous tissue and normal lung tissue in the same patient were analyzed. The methods were the same as that of Example 3, provided that b-actin gene was used as a control gene, and Pre-Developed TaqMan Assay Reagents Endogenous Control Human Beta-actin (Applied Biosystems Co.) was used in the quantification (Figure 2). Consequently, it was explained that both genes can be used at least as a lung cancer marker.

Example 5 Assay for acceptor substrates of glycosyl-transferase activities

For the reaction of GalNAc-T assay, 50 mM MES buffer (pH 6.5) containing 0.1 % triton X-100, 1 mM UDP-GalNAc, 10 mM $MnCl_2$ and 500 μM each acceptor substrate was used. A 10 μl of enzyme solution for 20 μl of each reaction mixture were added and incubated at 37 °C for various periods. After the incubation the mixture was filtrated with Ultrafree-MC column (Millipore, Bedford, MA), and 10 μl aliquot was subjected to reversed-phase high performance liquid chromatography (HPLC) on an ODS-80Ts QA column (4.6 x 250 mm; Tosoh, Tokyo, Japan). A 0.1 % TFA/H_2O with 12 % acetonitrile was used as a running solution. An ultraviolet spectrophotometer (absorbance at 210 nm), SPD-$10A_{VP}$ (Shimazu, Kyoto, Japan) was used for detection of the peaks. When the pyridyl amino-labeled oligosaccharides were utilized as acceptor substrates, 50 nM substrates were added into the reaction mixtures. For the analyses of the products derived from pyridyl amino labeled oligosaccharides, 100 mM acetic acid/triethylamine (pH4.0) was used as a running solution and the products were eluted with a 30-70% gradient of 1% 1-butanol in running solution at a flow rate of 1.0 ml/min at 55 °C.

A 200 μg of the reaction product was dissolved in 150 μl of D_2O using a micro cell and used as a sample for ^1H NMR experiments. One-dimensional and two-dimensional ^1H NMR spectra were recorded with DMX750 (Bruker, Germany, 750.13 MHz for ^1H nucleus) and ECA800 (JEOL, Tokyo, Japan, 800.14

MHz for ^1H nucleus) spectrometers at 25 °C. Methylene proton of benzyl group in higher field (4.576 ppm) was used as a reference for the ^1H NMR chemical shifts tentatively.

To investigate the specificity for acceptor substrates, N- and O-glycans containing GlcNAc on their non-reducing termini were utilized. As shown in Table 6 and 7, all acceptor substrates examined could receive a GalNAc residue.

Table 6

Substrate specificity of NGalNAc-Ts

Acceptor substrate	Relative activity (%)	
	NGalNAc-T1	NGalNAc-T2
1. GlcNAcβ-Bz	100	100
2. GlcNAcβ1-6(Galβ1-3)GalNAcα-*p*Np (core2-*p*Np)	15.2	11.4
3. GlcNAcβ1-3GalNAcα-*p*Np (core3-*p*Np)	20.0	32.3
4. GlcNAcβ1-6GalNAcα-*p*Np (core6-*p*Np)	190.7	220.4

Table 7

Substrate specificity of NGalNAc-Ts

	Acceptor substrate	Relative activity (%)	
		NGalNAc-T1	NGalNAc-T2
1.	GlcNAcβ1-2Manα1$\backslash_{6}$ GlcNAcβ1-2Manα1$/_{3}$Manβ1-4GlcNAcβ1-4 GlcNAc-PA	100	100
2.	Fucα1 GlcNAcβ1-2Manα1$\backslash_{6}$ GlcNAcβ1-2Manα1$/_{3}$Manβ1-4GlcNAcβ1-4 GlcNAc-PA	76.8	87.1
3.	Galβ1-4GlcNAcβ1-2Manα1$\backslash_{6}$ GlcNAcβ1-2Manα1$/_{3}$Manβ1-4GlcNAcβ1-4 GlcNAc-PA	26.2	45.0
4.	Fucα1 Galβ1-4GlcNAcβ1-2Manα1$\backslash_{6}$ GlcNAcβ1-2Manα1$/_{3}$Manβ1-4GlcNAcβ1-4 GlcNAc-PA	26.7	51.7
5.	GlcNAcβ1-2Manα1$\backslash_{6}$ Galβ1-4GlcNAcβ1-2Manα1$/_{3}$Manβ1-4GlcNAcβ1-4 GlcNAc-PA	16.2	21.6
6.	Fucα1 GlcNAcβ1-2Manα1$\backslash_{6}$ Galβ1-4GlcNAcβ1-2Manα1$/_{3}$Manβ1-4GlcNAcβ1-4 GlcNAc-PA	3.4	5.0

^{1}H NMR spectroscopy was performed to determine the newly formed glycosidic linkage of NGalNAc-T2 product. One-dimensional ^{1}H NMR spectrum of the NGalNAc-T2 product is shown in Fig. 5. In the NMR spectra, signal integrals (not shown, five phenyl protons of Bz, two methylene protons of Bz, two anomeric protons, twelve sugar protons except anomeric protons, six methyl protons of two N-acetyl groups) were in good correspondence with the structure of GalNAc-GlcNAc-O-Bz. As shown in Fig. 5 and in Table 8, two anomeric protons revealed resonances at very close magnetic field with coupling constant ($J_{1,2}$) larger than 8 Hz. This indicates that two pyranoses in the samples are in β-gluco-configuration. All ^{1}H signals could be assigned after high resolutional detections of COSY, TOCSY and NOESY

experiments. The anomeric resonance in the lower field showed NOE with two methylene protons of benzyl group in the sample (not shown), on the other hand, the anomeric resonance in higher field did not show NOE with methylene protons (not shown). The facts mean that the anomeric resonance in the lower field is responsible for the anomeric proton of the substrate pyranose (β-GlcNAc, defined as A), and that the anomeric proton in the higher field corresponds to anomeric proton of the transferred pyranose (β-GalNAc, defined as B). The chemical shifts and coupling constants of sugar part of the sample were shown in Table 8. The chemical shift and signal splitting of B-4 resonance was characteristic in β-Gal configuration [see Reference 15], and the order in chemical shift of A1-A6 protons was characteristically similar to observed spectrum of β-GlcNAc in LNnT (Galβ1-4GlcNAcβ1-3Galβ1-4Glc). As shown in Fig. 6, weak NOE cross peak between B1 and A4 and very weak NOE cross peaks between B1 and two A6 were observed in addition to strong inner residual NOEs between B1 and B5 and between A1 and A5. These suggest the existence of β1-4 linkage between two pyranoses. Results in NMR experiments thus indicated clearly that the product by NGalNAc-T2 is GalNAcβ1-4GlcNAc-O-Bz.

Table 8

Chemical shifts (ppm) and coupling constants (Hz) of sugar CH protons in the NGalNAc-T2 product

	NGalNAc-T2 product	
	GlcNAc	GalNAc
^{1}H Chemical shifts (*ppm*)[a]		
δ1	4.434	4.425
δ2	3.647	3.831
δ3	3.546	3.665
δ4	3.534	3.846
δ5	3.411	3.628
δ6	3.589	3.696
δ6	3.782	3.680
δCH_3	1.830	1.987
Coupling constants (*Hz*)		
$J_{1,2}$	8.5	8.4
$J_{2,3}$		10.8
$J_{4,5}$		<3.7
$J_{5,6a}$	5.6	<3.7
$J_{5,6b}$	2.0	
$J_{6a,6b}$	12.1	

a, The chemical shifts were set as the higher field signal of the benzyl methylene protons is ppm tentatively.

Example 6 LacdiNAc synthesizing activity of NGalNAc-T2 toward asialo/agalacto-fetal calf fetuin

As demonstrated in Table 6 and 7, both NGalNAc-T1 and -T2 transferred GalNAc toward both O- and N-glycans substrates. The LacdiNAc (GalNAcβ1-4GlcNAc) structures have been found in *N*-glycans of some glycoproteins in human. Therefore, to determine the activity of NGalNAc-T2 to transfer GalNAc to a glycoprotein, fetal calf fetuin (FCF), which has both N- and O-glycans, was utilized as an acceptor substrate.

Fetal calf fetuin (FCF), neuraminidase, β1-4 galactosidase and glycopeptidase F were purchased from Sigma, Nacalai Tesque (Kyoto, Japan), Calbiochem and Takara, respectively. Asialo/agalacto-FCF was prepared from 200 μg of FCF by incubating with 4 μU of neuraminidase and 12 μU of β1,4-galactosidase at 37 °C for 16 hr. The transfer of GalNAc by GalNAc-T2 to glycoprotein was performed in 20 μl of a standard reaction mixture containing 50 μg of asialo/agalacto-FCF produced by glycosidase treatment. After the incubation at 37°C for 16 hr, each 5 μl of the reaction mixture was digested with glycopeptidase F (GPF) according to manufacture's instruction. For detection of transferred GalNAc, horseradish peroxidase (HRP) conjugated lectin, Wisteria floribunda agglutinin (WFA) (EY Laboratories, San Mateo, CA), was used. A 1 μl of reaction mixtures subjected to 12.5% SDS-PAGE were transferred to nitrocellulose membrane (Schleicher & Schuell, Keene, NH) and stained with 0.1% HRP conjugated WFA lectin. The signals were detected using enhanced chemiluminescence (ECL) and Hyperfilm ECL (Amersham Biosciences).

As shown in Fig. 3, asialo/agalacto-FCF appeared as approximately 55 and 60 kDa band (lane 1). NGalNAc-T2 effectively transferred GalNAc to asialo/agalacto-FCF (lane 5). Furthermore, the band mostly disappeared by a GPF treatment, and its molecular size was detected at approximately 45 and 50 kDa position by Coomassie staining (Fig. 3, lane 3 and 6). In the case of NGalNAc-T1, the activity toward asialo/agalacto-FCF was same as NGalNAc-T2 (data not shown).

Example 7 Analysis of N-glycan structures on glycodelin from NGalNAc-T1 and -T2 gene transfected CHO cells

As shown above, both NGalNAc-T1 and -T2 could synthesize LacdiNAc structures on mono- and oligosaccharide acceptors. Actually, it is known that the LacdiNAc structures exist in N-glycans on some glycoproteins. Therefore we examined the ability of NGalNAc-T1 to construct LacdiNAc on glycodelin, which is one of major glycoproteins carrying LacdiNAc structures, *in vivo*. CHO cells were employed for this purpose, because glycodelin produced in CHO cells is devoid of any of the LacdiNAc-based chains.

The glycodelin expression vector was transfected into CHO cells expressing NGalNAc-T1 or -T2 gene and the culture medium was collected from 48 hr-culture medium. Glycodelin was harvested with WFA affinity column from the culture medium. The harvested glycodelin was applied to SDS-PAGE and used for lectin blotting with WFA.

As shown in Fig. 7, the non-reducing terminal GalNAc was detected only when NGalNAc-T1 or -T2 gene was co-transfected with glycodelin gene. These bands were disappeared by N-glycanase™ treatment, therefore these GalNAc residues might exist in N-glycans.

Example 8 Preparation of mouse proteins of the present invention

1. Search through a genetic database and determination of

the nucleic acid sequence of a novel mouse N-acetylgalactosaminyltransferase

A search of similar genes through a mouse genomic database (UCSC Human Genome Project, Nov. 2001 mouse assembly archived Sep. 15, 2002, http://genome-archive.cse.ucsc.edu/) was performed by use of the genes for existing human NGalNAc-T1 and -T2. The sequences used were SEQ ID NOs: 1, 3, 26 and 28. The search was performed using a program such as Blast [Altschul et al., J. Mol. Biol., 215, 403-410 (1990)].

As a result, two homologous genes were found on mouse chromosome 7 and 6. The nucleotide and amino acid sequences of the first gene on chromosome 7, which is an ortholog of human NGalNAc-T1, were shown as SEQ ID NOs: 26 and 28. The second ones on chromosome 6 were described as SEQ ID NOs: 27 and 29.

2. Integration of GalNAc-T genes into an expression vector

To prepare each expression system of mouse NGalNAc-T, a portion of each gene was first integrated into pFLAG-CMV1 (Sigma).

Integration of mNGalNAc-T1 into pFLAQ-CMAV1

The mouse NGalNAc-T2 (mNGalNAc-T2) gene encoding its putative catalytic domain (amino acid 45 to 1,034) was amplified with two primers, 5'-CCC AAG CTT CGC CTG GGC TAC GGG CGA GAT-3' (SEQ ID NO: 31) and 5'-GCT CTA GAC TCA GGA TCG CTG TGC GCG GGC A-3' (SEQ ID NO: 32), using the cDNA derived from mouse brain as a template. The mRNA was

prepared from mouse brain with RNeasy mini kit (Qiagen), then the cDNA was synthesized with SuperScript first-strand synthesis system for RT-PCR (Invitrogen). For the PCR, LA Taq DNA polymerase (Takara) was used. The amplified 2.7 kb fragment was digested with endonuclease Hind III and Xba I, then the digested fragment was inserted into pFLAG-CMV-1 and pFLAG-mNGalNAc-T1 was constructed.

Integration of mNGalNAc-T2 into pFLAG-CMAV1

The mouse NGalNAc-T2 (mNGalNAc-T2) gene encoding its putative catalytic domain (amino acid 57 to 986) was amplified with two primers, 5'-CCC AAG CTT CGG CCC AGG CCG GCG GGA ACC-3' (SEQ ID NO: 33) and 5'-GGA ATT CTC ACG GCA TCT TCA TTT GGC GA-3' (SEQ ID NO: 34), using the cDNA derived from mouse stomach as a template. The mRNA was prepared from mouse stomach with RNeasy mini kit (Qiagen), then the cDNA was synthesized with SuperScript first-strand synthesis system for RT-PCR (Invitrogen). For the PCR, LA Taq DNA polymerase (Takara) was used. The amplified 2.7 kb fragment was digested with endonuclease Hind III and EcoR I, then the digested fragment was inserted into pFLAG-CMV-1 and pFLAG-mNGalNAc-T2 was constructed.

3. Transfection and expression of recombinant enzymes

A 15 μg of pFLAG-mNGalNAc-T1 or pFLAG-mNGalNAc-T2 was induced into 2 X 10^6 of HEK293T cells which were cultured overnight in DMEM (Dulbecco's modified Eagle's medium) including 10 % FCS (fetal calf serum), using Lipofectamine 2000 (Invitrogen Co.) as a protocol provided

by the same company. A supernatant of 48-72 hors was collected. The supernatant was mixed with NaN_3 (0.05 %), NaCl (150 mM, $CaCl_2$ (2 mM) and an anti-M1 resin (Sigma Co.) (50 µl), and the mixture was stirred overnight (3000 rpm, 5 min, 4 °C) to collect a pellet. The pellet was combined with 900 µl of 2 mM $CaCl_2$/TBS and re-centrifuged (2000 rpm, 5 min, 4 °C), after which the pellet was suspended in 200 µl of 1 mM $CaCl_2$/TBS to give a sample for assaying activity (mNGalNAc-T1 or mNGalNAc-T2 enzyme solution).

The enzyme was subjected to conventional SDS-PAGE and Western blotting, and the expression of the intended protein was confirmed. Anti-FLAG M2-peroxydase (A-8592, SIGAIA Co.) was used as an antibody.

References

1. Sugita, M., S. Itonori, F. Inagaki and T. Hori, Characterization of two glucuronic acid-containing glycosphingolipids in larvae of the green-bottle fly, Lucilia caesar. J. Biol. Chem., 1989. 264, p.15028-33

2. Helling, F., R.D. Dennis, B. Weske, G. Nores, J. Peter-Katalinic, U. Dabrowsli, H. Egge and H. Wiegandt, Glycosphingolipids in insects. The amphoteric moiety, N-acetylglucosamine-linked phosphoethanolamine, distinguishes a group of ceramide oligosaccharides from the pupae of Calliphora vicina (Insecta: Diptera). Eur. J. Biochem., 1991. 200, p.409-21

3. Weisshaar, G., J. Hiyama, A.G. Renwick and M. Nimtz, NMR investigations of the N-linked oligosaccharides at individual glyocosylation sites of human lutropin. Eur. J.

Biochem., 1991. 195, p.257-68

4. White, T., E.P. Bennet, K. Takio, T. Sorenesen, N. Bonding and H. Clausen, Purification and cDNA cloning of a human UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase. J. Biol. Chem., 1995. 270, p.24156-65

5. Bennett, E.P., H. Hassan and H. Clausen, cDNA cloning and expression of a novel human UDP-N-acetyl-alpha-D-galactosamine. Polypeptide N-acetylgalactosaminyltransferase, GalNAc-t3. J. Biol. Chem., 1996. 271, p.17006-12

6. Bennett, E.P., H. Hassan, U. Mandel, E. Mirgorodskaya, P. Roepstorff, J. Burchell, J. Taylor-Papadimitriou, M.A. Hollingsworth, G. Merkx, A.G. van Kessel, H. Eiberg, R. Steffensen and H. Clausen, Cloning of a human UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase that complements other GalNAc-transferases in complete O-glycosylation of the MUC1 tandem repeat. J. Biol. Chem., 1998. 273, p.30472-81

7. Bennett, E.P., H. Hassan, J. Mandel, M.A. Hollingsworth, N. Akisawa, Y. Ikematsu, G. Merkx, A.G. van Kessel, S. Olofsson and H. Clausen, Cloning and characterization of a close homologue of human UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase-T3, designed GalNAc-T6. Evidence for genetic but not functional redundancy. J. Biol. Chem., 1999. 274, p.25362-70

8. Bennett, E.P., H. Hassan, M.A. Hollingsworth and H. Clausen, A novel human UDP-N-acetyl-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase,

GalNAc-T7, with specificity for partial GalNAc-glycosylated acceptor substrates. FEBS Lett., 1999. 460, p.226-30

9. White, K. E., B. Lorenz, T. Meitinger, T.M. Strom and M.J. Econs, Gene, 2000. 246, p.347-56

10. Toba, S., M. Tenno, M. Konishi, T. Mikami, N. Itoh and A. Kurosaka, Brain-specific expression of a novel human UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase (GalNAc-T9). Biochim. Biophys. Acta., 2000. 7, p.264-8

11. Nagata, Y., S. Yamashiro, J. Yodoi, K.O. Lloyd, H. Shiku and K. Furukawa, Expression cloning of beta 1,4 N-acetylgalactosaminyltransferase cDNAs that determine the expression of GM2 and GD2 gangliosides. J. Biol. Chem., 1992. 269, p.12082-9

12. Yamamoto, F., J. Marken, T. Tsuji, T. White, H. Clausen and S. Hakomori, Cloning and characterization of DNA complementary to human UDP-GalNAc:Fuc alpha 1----2Gal alpha 1----3GalNAc transferase (histo-blood group A transferase) mRNA. J. Biol. Chem., 1990. 265, p.1146-51

13. Xu, H., T. Storch, M. Yu, S.P. Elliott and D.B. Haslam, Characterization of the human Forssman synthetase gene. An evolving association between glycolipid synthesis and host-microbial interactions. J. Biol. Chem., 1999. 274, p.29390-8

14. Guo, J.M., Y. Zhang, L. Cheng, H. Iwasaki, H. Wang, T. Kubota, K. Tachibana and H. Narimatsu, Molecular cloning and characterization of a novel member of the UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase family, pp-GalNAc-T12(1). FEBS Lett., 2002. 524, p.211-8

15. Ishizuka, Y., T. Nemoto, M. Fujiwara, K. Fujita and H. Nakanish, Three-dimensional structure of fucosyllactoses in

an aqueous solution. J. Carbohydr. Chem., 1999. 18, p.523-33

Industrial Applicability

According to the present invention, an enzyme which transfers N-acetylgalactosamine to N-acetylglucosamine via a β1-4 linkage was isolated and the structure of its gene was explained. This led to the production of said enzyme or the like by genetic engineering techniques, the production of oligosaccharides using said enzyme, and the diagnosis of diseases on the basis of said gene or the like.